Canonical Correlation Using For Studying The Personal Factors Trace in Using Mobile

Ehab A. Mahmood

Administration & Economic College, Babylon University

Abstract

The Mobile device is very important in our days life, and there is many uses for it according to user's personality, Where the continuous technical development leads to multitude uses in several purposes , so in this article it will study the most important personal factors and ranges of these effects in how the mobile can use and this realizes in using canonical correlation style to sample of size (150) including male and female ,by giving questionnaire form to then which includes 11(6) questions relate with personality of person and (25) questions relate with the mobile using to determine which the variables that is more effect in the using it .

الخلاصة

ان لجهاز الموبايل اهمية واسعة في حياتنا اليومية وتتعدد استخداماته تبعا لشخصية الفرد ، حيـث ان التطـور التقنـي المستمر ادى لتعدد استخداماته في اغراض عديدة ، لذا سيتم في هذا البحث دراسة اهم العوامل الشخصية ومدى تأثيرها في كيفية استخدام الموبايل باتباع اسلوب الارتباط القويم لعينة بحجم (١٥٠) شملت ذكورا وإناثا ، من خلال توزيع استمارة الاستبيان عليهم والتي شملت (٦) اسئلة متعلقة بشخصية الفرد و(٢٥) سؤال متعلق باستخدام الموبايل لتحديد أي المتغيـرات الكثـر تـأثيرا فـي استخدامه.

١-١ Introduction

The Canonical Correlation Analysis (CCA) is one of the most important style in multivariate statistical analysis, in case of studying two groups of variables where , the first is independent variable and the second is dependent variable. In this paper it will depended on the procedure of (CCA) to study the effect of the personal factors as it is independent variables for the mobile using and the variables which relate with the how mobile is using as dependent variables, and determines which the trace is more active for both male and female.

Hotelling (1936) reviewed the method of Canonical Correlation Analysis and the other researcher had been his method as :

Dehon and others (2000) compared four ways for performing a robust (CCA), one method used robust estimators of the involved covariance matrices, second used the signs of the observations, third was based on projection pursuit, and finally an alternative regression algorithm for canonical analysis. They applied two experiments: the first generated data for $N(0,\Sigma)$ distribution, and the second experiment were generating the data as before, but added (10%) contamination data from $N(0,50I_p)$, and they said that on the clean data the classical (CCA) was the most precise , and it looses its optimality in the contaminated case where it is outperformed by the robust methods .

Borga & Knutsson (2001) presented a method for learning visual operator by (CCA) and they revealed the experimental results which indicated that the proposed method is applicable on a wide range of different visual tasks .

Hendrix and others (2001) showed that the (CCA) was considerable promise as the foundation of a multivariate graphical data analysis method for two or more linked multivariate datasets .

Lai (2002) derived an artificial neural network as an implementation of the standard statistical technique of (CCA) , and he compared its capability with standard statistics methods on both real and artificial data .

Weenink (2003) discussed algorithms for performing (CCA) , and his results can be used with modern robust statistics against outliers.

Huang and others (2006) studied nonlinear association measures by using Kernel method , where the Kernel Canonical Correlation Analysis (KCCA) is a method that generalizes the classical linear (CCA) to nonlinear setting .

Via and others (2006) studied the (CCA) and introduce algorithms for multiple data set (more than two data set) .

Gonzalez and others (2008) propose an efficient way to perform (CCA) in (R); the functions provide in the (CCA) package outperform the cancor (R) function according to several points :

* The ability to handle missing values .
* The processing of data sets with more variables than units through the regularized extension of the (CCA) .
* Integrated solutions for graphical outputs .

Vaerenbergh and others (2008) propose Adaptive (KCCA) for the nonlinear system .

Kakade and Foster (2008) study multi – view regression problem via (CCA) , and display number of lemmas and their proofs .

١ -2 Objective

It is the studying the effects of the important personal factors which relate with user of the mobile and determine which the factors are more effect and determine which the variables that relate with the mobile using are more responding to these factors.

2 – 1 Key - Terms - Definition : [Hair & others (1998)]

1-Canonical Variates :

They are the linear combinations that represent the total weighted of two or more variables and can be defined either independent or dependent variables , also referred to linear composites or linear compounds .

2 - Canonical Function :

It is the relationship between linear composites (canonical variates).

Each Canonical function has two Canonical variates, first for the group of independent variables and the second for the group of dependent variables . The strength of the relationship is given by the Canonical Correlation.

3 - Canonical Roots :

They are the squared of canonical correlations , which estimate the amount of shared variance between the respective optimally weighted canonical variates independent and dependent variables . Also known as (Eigenvalues).

2 - 2 Canonical Correlation Analysis (CCA) :

There are several basic ideas about the (CCA) such as :

* It is a multivariate statistical model that facilitates the study of interrelationships among sets of multiple dependent variables and multiple independent variables [Green(1978)' Green and Carroll (1978)].
* The (CCA) cannot be performed when the sample size is less than the greatest amount of variables in both data set. {Gonzalez and others (2008)}.
* The (CCA)is a classical tool in statistical analysis that measures the linear relationship between two or several data sets. Via and others(2006),Huang and others(2006)]. Where as Principal Components Analysis and Factor Analysis deal with the inter relationship within a set of variables. [Lai(2002)]

The (CCA) originates in Hotelling (1936) and the two equations that govern the analysis are the following : [Weenink(2003)]

$$(\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} - \rho^2\Sigma_{yy})\mathrm{Y} = 0$$
$$(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} - \rho^2\Sigma_{xx})\mathrm{X} = 0 \qquad \ldots\ldots\ldots\ldots (1)$$

Where: Σ'_{xy} denotes the transpose of Σ_{xy} .

2 - 3 Derivation of The (CCA) Equations : [Weenink(2003)]

Let A_x and A_y be the two data sets have the dimensions (m*n) and (m*p) respectively , and the maximum number of correlations that we can find is that equal to the minimum of the column dimensions (n) and (p) .

Let the directions of optimal correlations for the (A_x) and (A_y) data sets be given by the vectors (X) and (Y) respectively ,When it project our data on these direction vectors , we obtain two new vectors (ξ_x) and (ξ_y) defined as follows:

$$\xi_x = \mathrm{A}_x\mathrm{X} \qquad \ldots\ldots\ldots\ldots (2)$$
$$\xi_y = \mathrm{A}_y\mathrm{Y} \qquad \ldots\ldots\ldots\ldots (3)$$

The variables (ξ_x) and (ξ_y) are called the Canonical Variates and the correlation between them given by :

$$\rho = \frac{\xi'_x\xi_y}{\sqrt{(\xi'_x\xi_x)(\xi'_y\xi_y)}} \quad (\rho) \qquad \ldots\ldots\ldots\ldots (4)$$

To find the directions (x) and (y) that maximize (ρ) ,we shall assume that the :

$$\xi'_x\xi_x = \mathrm{X}'\mathrm{A}'_x\mathrm{A}_x\mathrm{X} = \mathrm{X}'\Sigma_{xx}\mathrm{X} = 1 \qquad \ldots\ldots\ldots\ldots (5)$$
$$\xi'_y\xi_y = \mathrm{Y}'\mathrm{A}'_y\mathrm{A}_y\mathrm{Y} = \mathrm{Y}'\Sigma_{yy}\mathrm{Y} = 1 \qquad \ldots\ldots\ldots\ldots (6)$$

And we get by the Lagrangian form :

$$L(\rho_x, \rho_y, \mathrm{X}, \mathrm{Y}) = \mathrm{Y}'\Sigma_{yx}\mathrm{X} - \frac{\rho_x}{2}(\mathrm{X}'\Sigma_{xx}\mathrm{X} - 1) - \frac{\rho_y}{2}(\mathrm{Y}'\Sigma_{yy}\mathrm{Y} - 1) \qquad \ldots\ldots\ldots\ldots (7)$$

Where : Σ'S are covariance matrices .

The first derivatives with respect to X and Y :

$$\frac{\partial L}{\partial X} = \Sigma_{xy}\mathrm{Y} - \rho_x\Sigma_{xx}\mathrm{X} = 0 \qquad \ldots\ldots\ldots\ldots (8)$$
$$\frac{\partial L}{\partial \mathrm{Y}} = \Sigma_{yx}\mathrm{X} - \rho_y\Sigma_{yy}\mathrm{Y} = 0 \qquad \ldots\ldots\ldots\ldots (9)$$

By the subtract (X') times the first equation from (Y') times the second and we have :

$$\mathrm{Y}'\Sigma_{yx}\mathrm{X} - \rho_y\mathrm{Y}'\Sigma_{yy}\mathrm{Y} - \mathrm{X}'\Sigma_{xy}\mathrm{Y} + \rho_x\mathrm{X}'\Sigma_{xx}\mathrm{X} = 0 \qquad \ldots\ldots\ldots\ldots (10)$$
$$\therefore \rho_x\mathrm{X}'\Sigma_{xx}\mathrm{X} - \rho_y\mathrm{Y}'\Sigma_{yy}\mathrm{Y} = 0 \qquad \ldots\ldots\ldots\ldots (11)$$

By substitute equations (5) and (6) in equation (11) we get :

$$\rho_x - \rho_y = 0 \qquad \ldots\ldots\ldots\ldots (12)$$
$$\rho_x = \rho_y = \rho \qquad \ldots\ldots\ldots\ldots (13)$$

From equations (8) and (9) we get :

$$\mathrm{X} = \frac{\Sigma_{xx}^{-1}\Sigma_{xy}\mathrm{Y}}{\rho} \qquad \ldots\ldots\ldots\ldots (14)$$
$$\mathrm{Y} = \frac{\Sigma_{yy}^{-1}\Sigma_{yx}\mathrm{X}}{\rho} \qquad \ldots\ldots\ldots\ldots (15)$$

By substitute equations (14) in equation (9) and equation (15) in equation (8) we get :

$$\Sigma_{xy}\frac{\Sigma_{yy}^{-1}\Sigma_{yx}X}{\rho}-\rho\Sigma_{xx}X=0 \qquad \ldots\ldots\ldots\ldots\ldots (16)$$

$$\Sigma_{yx}\frac{\Sigma_{xx}^{-1}\Sigma_{xy}Y}{\rho}-\rho\Sigma_{yy}Y=0 \qquad \ldots\ldots\ldots\ldots\ldots (17)$$

Then we get :

$$(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}-\rho^2\Sigma_{xx})X=0 \qquad \ldots\ldots\ldots\ldots\ldots (18)$$

$$(\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}-\rho^2\Sigma_{yy})Y=0 \qquad \ldots\ldots\ldots\ldots\ldots (19)$$

Because the matrices (Σ_{xy}) and (Σ_{yx}) are each other's transpose we write the canonical correlation analysis equations as follows :

$$(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy}-\rho^2\Sigma_{xx})X=0 \qquad \ldots\ldots\ldots\ldots (20)$$

$$(\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}-\rho^2\Sigma_{yy})Y=0 \qquad \ldots\ldots\ldots\ldots (21)$$

3 - 1 Data Description :

It will distribute questionnaire form which is illustrative in the appendix for sample size (150) includes (75) male and (75) female , have two variables groups , the first group : Represent (6) independent variables which relate to the person himself as (Sex, Age , Educational Level , Social State , Home , Income).

The second group represents (25) dependent variables which relate with how the mobile device is using .

3 - 2 Results Analysis :

By using the statistica program it will observe that's :

It is from the table (1) , observing that the value of the first canonical correlation is (0.7409) and the (χ^2) value is (297.72) and (p= 0) with degree of freedom is (150), that Leads to the correlation between the two groups is very significant for all ranges , it notes that the value of eigenvalue is (0.5489), that means that the first canonical variable for the first group illustrate is (54.89 %) from the variance of canonical variable which is identical for the two group .

The value of the second canonical correlation is (0.6569) and the value of (χ^2) is (191.84) and (p=0.000036) with degree of freedom (120) , so this leads to the correlation between the two groups is significant for all level , and it notes that the second eigenvalue is (0.4316), that means that the second canonical variable for the first group illustrate (43.16%) from the variance of canonical variable which is identical for the two group .

Also the value of the third canonical correlation is (0.5036) and the recorded of (χ^2) is (116.71) and (p=0.0423) with degree of freedom (92) , that leads to the correlation between two groups is significant , and the value of the third eigenvalue is (0.2536) , that means that the illustrate ratio is (25.36 %) .

The another canonical correlations are non significant with the degree of freedom which is recorded for each one of them .

The plot (1) illustrates the canonical correlation , and the plot (2) illustrates the eigenvalue .

It will illustrate the results according to the ordering of the roots which find in table (2) and (3) as following :

3-2-1 Root (1) :

It is showing the sex variable is more effecting for the group of dependent variables and it is positive effect, then educational level but with opposite effect , then income variable with positive effect, and the other variables have low effects .

And for the group of dependent variables, it is observing that the question (22) is the more one which effects on the group of independent variables and it is positive response, then the question (24) which is also positive effect, then the question (10) but in opposite response ,then the question (20) with positive response, then the question (15) with opposite response, then question (19) is opposite effect also , then the questions (14) and (13) are positive . Effect also about the other variables response is very low .

3-2-2 Root (2) :

In the group of independent variables it is observing that the social state is more effective in the group of dependent variables, then the age variable and sex variable and they are all have the opposite effect, then educational level with positive effect , about the others variables have the low effect .

For the group of dependent variables it is noting that the question (15) is more response for the independent variables and this response is positive , then the question (1) is positive response also , then the questions (23) then (17) then (12) all have opposite response , then the questions (9) then (3) that have positive response , and the other variables have low response in general .

3-2-3 Root (3) :

In the group of independent variables it is observing that the age variable is more effect for the dependent variables which have opposite effect , then social state variables but this effect is positive , then home variable with positive effect also , then income variable but with the opposite effect , then educational level which has positive effect .

For the second group of variables the question (13) is more one which is responding to the group of independent variables which is positive response , then the questions (1) , (16) , then (6) is opposite response , then the question (5) is positive response , then the questions (11) then (24) which have opposite response , then the questions (9) then (22) with positive response .

Table (1) Chi-Square Tests with Successive Roots Removed (new.sta)

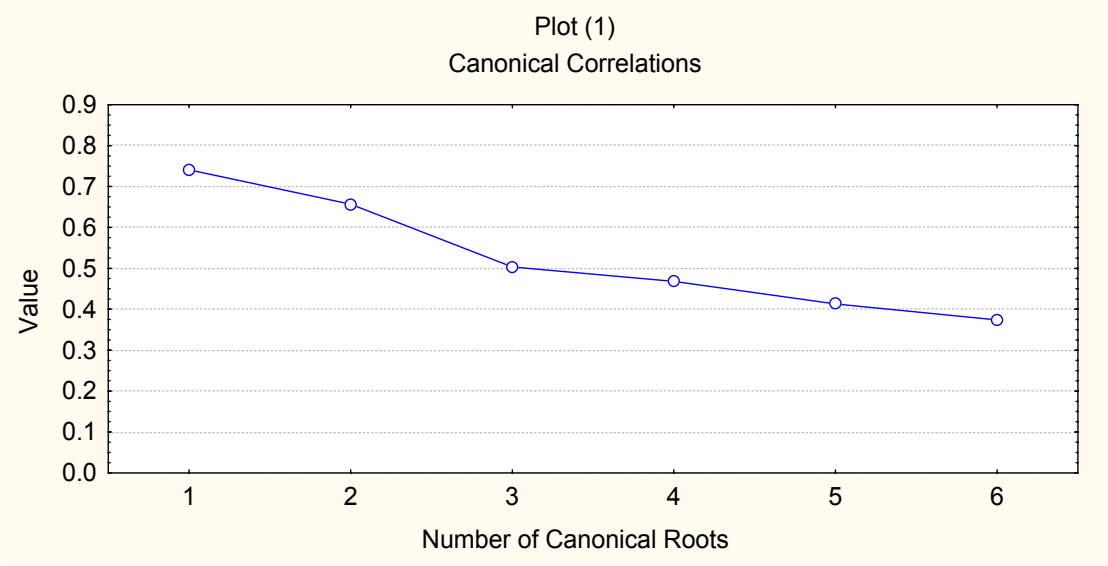
Root Removed	Canonical (R)	Canonical (R) – Sqr.	Chi– Sqr .	df	P	Lambda Prime
0	.740887	.548914	297.7187	150	.000000	.106620
1	.656944	.431575	191.8378	120	.000036	.236362
2	.503627	.253640	116.7081	92	.042253	.415819
3	.468211	.219221	77.7993	66	.152111	.557130
4	.412969	.170543	44.8867	42	.351943	.713556
5	.373805	.139730	20.0177	20	.456860	.860270

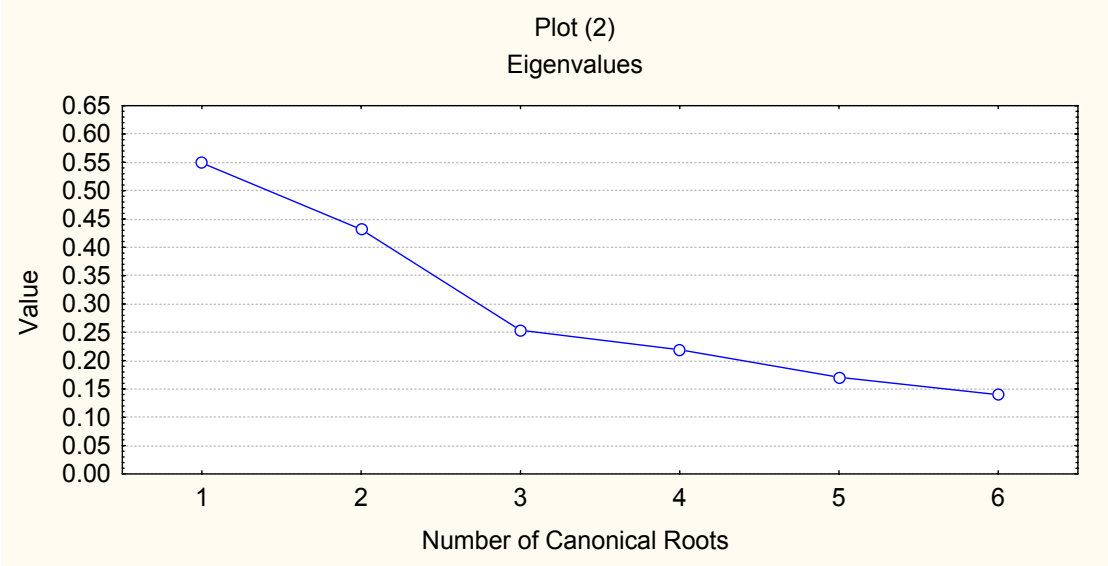
Table (2)
Canonical Weights, left set (new.sta)

Root	Root 1	Root 2	Root 3	Root 4	Root 5	Root 6
Sex	.785507	-.270877	.15310	.447239	.412407	.042377
Age	-.051917	-.399780	-1.43244	-.295275	.505359	.502759
Social State	-.028139	-.672095	1.13102	-.419536	-.755940	-.077078
Educational Level	-.531141	.133377	.30671	1.016727	.336681	-.472529
Home	.086608	-.028779	.41988	-.444995	.765371	-.356003
Salary	.376445	-.045349	-.40037	-.033229	-.364183	-.781832

Table (3) Canonical Weights, right set (new.sta)

Root	Root 1	Root 2	Root 3	Root 4	Root 5	Root 6
Q1	-.117868	.347848	-.475350	-.134106	-.255904	-.265019
Q2	-.063308	-.142856	.067818	-.109267	.334412	.083000
Q3	.093356	.202600	.004662	.411865	-.028597	.326989
Q4	.097358	.057241	-.181936	-.153679	.002945	-.144617
Q5	-.173764	-.136591	.371258	-.099926	.144385	.209098
Q6	.137694	-.003393	-.388129	.208650	-.021979	.124525
Q7	.166762	.126748	.144720	.129290	-.146428	-.131203
Q8	-.040788	.132734	-.000423	.424508	-.165284	.165407
Q9	.045268	.204270	.339793	-.122269	-.239613	-.020362
Q10	-.298574	.110539	-.073039	-.080394	.405997	-.116765
Q11	.010320	.140858	-.364025	.013960	-.120728	.055207
Q12	-.048399	-.214107	-.102445	-.049508	.062135	-.665365
Q13	.202597	.017561	.544680	-.293222	.288486	.117356
Q14	.259998	.143353	.197304	-.124058	-.592826	.384585
Q15	-.260909	.535691	.137810	-.096949	.452328	-.016827
Q16	.037555	-.142246	-.410122	-.026558	.367489	-.021344
Q17	.121739	-.226254	.079374	.101943	.019838	-.046424
Q18	-.157991	-.050944	.298470	.014769	-.344316	-.265601
Q19	-.259478	.032679	.125307	.287099	.222555	.278615
Q20	.296783	-.111244	-.287273	-.056965	.101173	.245653
Q21	.188397	.039542	-.004535	.076265	.399495	-.200110
Q22	.432645	.054437	.329100	.011756	-.087878	-.002460
Q23	.169018	-.235591	.107416	.622916	.188652	.010248
Q24	.378483	.012537	-.348423	.009008	-.027530	-.431644
Q25	-.073676	.188747	.043891	.406803	-.004206	-.134712





4 – 1 Conclusions

According to the background analysis in generally it is concluding the following :

1- The sex variable is more independent variables is effect in the dependent variables group as positive effect , then educational level with opposite effect , then the income variable with positive effect .

2- For the third root it appears clearly that the age variable is opposite effect and also social state variable but with the positive effect .

3- For the group of dependent variables , it is observed that the question (22) has more response for the group of independent variables as positive response ,then the question (24) with the positive response also , then the question (10) but the opposite response , then the question (20) with the positive response , then the question (15) with the opposite response , then the question (14) has positive response .

4- For the second root it is notes that the question (15) and (1) are more response for the group of independent variables with the positive response .

5- For the third root it notes that the question (13) is the more positive response , but the question (1) is opposite .

4 – 2 Recommendations :

1- Add new variables to the two groups and replay the application .
2- Using another statistical procedure for analyzing .

Appendix :
Questionnaire Form

Sex : Male – Female .
Age : (Under 20) – (20 – 25) – (25 – 30) – (30 – 35) – (35 – 40) – (40 and More) .
Social State : Married – Unmarried .
Educational Level : (Lower of Secondary) – (Secondary) – (Diploma) – (Bsc) - (Msc & PhD) .
Home : City Centre – Village .
Income : (Under 250 000) – (250 000 – 500 000) – (500 000 – 750 000) – (750 000 – Million) – (Million and More)

Question	Agreement Ratio				
	0%	25%	50%	75%	100%
1- Is the message service useful .					
2 – Is there any easiness and speedness in calling any time .					
3 – Are you using the mobile as a device to assuring for others especialy in this time .					
4 – Are you using it to exchange congratulations or personal comforting					
5- Are you using it to exchange congratulations or religious comforting .					
6- Is there any benefit from the internet service in the mobile .					
7 - Is there any benefit from the continuous technical development for the devices and servises increasing .					
8 - Is the mobile useful to know the date and time .					
9- Are you using it as alarm for waking .					
10 - Are you using it as remembrance for some an important occasion in your life .					
11 - Are you using the games in your mobile .					
12 - Are you using it as acamera device					
Question	Agreement Ratio				
	0%	25%	50%	75%	100%
13 - Are you saving Artists photo in your mobile .					
14 - Are you saving players photo in your mobile .					
15 - Are you hearing the songs in your mobile .					
16 - Are you seting up some scientific					

programs in your mobile .					
17 – Are you consider it one of the devices that develop children mentality.					
18 – Are you plug your mobile to computer or TV to play the recording .					
19 – Are the recharge of card mobile is effecting on your economically .					
20 – Are you using it to call the another sex .					
21 – Are you turn off it in some place as (class , meeting , sad occasions , els) .					
22 – Are you consider it as a device for threatening by unknown persons .					
23 – Are you using it when you drive the car .					
24 – Are you calling with others at late time .					
25 – Are you believe that the mobile has bad effect on health .					

References

Borga, M. & Knutsson, H.(2001). "Canonical Correlation Analysis in Early Vision Processing" .

Dehon, C., Filzmoser, P. and Croux, C. (2000). “Robust Methods for Canonical Correlation Analysis “ .

Gonzalez , I., Dejean, S., Martin, P.G.P. and Baccini, A.(2008). "CCA: An R PacKage to Extend Canonical Correlation Analysis". Journal of Statistical Software,Volume (23),Issue (12).

Green, P.E. (1978) . "Analyzing Multivariate Data". Hinsdale, I11: Holt , Rinehart & Winston .

Green, P.E. & Caroll, J.D. (1978)."Mathematical Tools for Applied Multivariate Analysis " New York , Academic Press .

Hair, J.F., Jr, Anderson ,R.E., Tatham, R.L. and Black, W.C.(1998). "Multivariate Data Analysis " 5th edition ,Prentice Hall.

Hendrix, K.A., Brown, B.L. and Hendrix, S.B. (2001) .”Canonical Correlation : The Underused Method“ .Joint Statistical Meetings–Statistical Computing Section . (1430 – 1434) .

Hotelling ,H. (1936)." Relations between two sets of variates". Biometrika, 28,(321-377).

Huang, S., Lee, M. and Hsiao, C.K. (2006). "Kernel Canonical Correlation Analysis and its Applications to Nonlinear Measures of Association and test of Independence".

Kakade, S.M. and Foster, D.P. (2008) “Multi–View Regression Via Canonical Correlation Analysis “.

Lai, P.L. (2002). "Probabilistic Derivation and Multiple Canonical Correlation Analysis " .

Vaerenbergh, S.V., Via, J. and Santamaria, I. (2008)."Adaptive Kernel Canonical Correlation Analysis Algorithms for Nonparametric Identification of Wiener and Hammerstein Systems" .

Via, J., Santamaria ,I. and Perez, J. (2006)."Canonical Correlation Analysis (CCA) Algorithms For Multiple Data Sets: Application to Blind Simo Equalization".

Weenink, D. (2003). "Canonical Correlation Analysis" Proceedings (25), (81- 99) .